

Iris Center Localization Using Energy Map Synthesis Based on Gradient and Isophote

Lihong Dai , Jinguo Liu , Zhaojie Ju and Yang Gao

Abstract. Gaze tracking has wide applications such as in driver fatigue detection, virtual reality, and human-computer interaction. The performance of gaze tracking depends largely on the accuracy of iris center localization. However, most of the existing gaze tracking products are intrusive or require additional equipment with a high cost. Therefore, precise localization methods of iris center in low quality images captured in a non-contact way with visible light need to be investigated. This paper proposes a novel localization method of iris center using energy map synthesis based on image gradient, isophote and midpoint of eye ROI (Region of interest). This method combines the advantages of higher localization accuracy based on gradient, invariance to the rotation and linear transformation of light based on isophote, and iris center close to the midpoint of eye ROI. Moreover, a post-processing correction method for the closed eyes and for other large deviations of iris center position is adopted to further improve the localization accuracy. The algorithm is verified on the BioID, Talking Face Video and MUCT Face databases, and the results show that the localization accuracy has outperformed the listed state-of-the-art methods in varying illuminations.

Keywords: Iris center localization, energy map synthesis, gradient, isophote, post-processing correction

1. Introduction

1.1. Significance of iris center localization

Gaze tracking has wide applications such as in driver fatigue detection[21], diagnosis of eye diseases and psychological diseases[2], website or advertising design, virtual reality and augmented reality[11], human-computer interaction[13], control equipment, and human behavior research. The performance of gaze tracking depends largely on the accuracy of iris center localization.

However, the popular gaze tracking products often require high-resolution cameras, infrared light sources, and other additional equipment. They are not only intrusive, but also expensive. Furthermore, images are often subjected to varying illuminations, specular reflection, occlusion, and so on, so it is difficult to accurately locate the iris centers of these images. Therefore, accurate localization methods of iris center in low quality images captured in a non-contact way with visible light need to be investigated.

1.2. Existing iris center localization methods

According to the form of light source, the iris center localization methods can be divided into those with active light source[25] and those with passive light source. For the former, additional infrared equipment is required, which is invalid in sunlight and long distances and usually only suitable for indoor use. While the latter is mostly adopted to locate the iris center, that is, the localization is carried out in the visible light environment. For the localization methods with passive light source, which can also be divided into the appearance-based methods and the feature-based methods. In the appearance-based methods, supervised machine learning is mainly used to locate the iris center. Markuš et al. adopt a supervised learning method of ensemble of regression tree to locate the iris center [12]. Since the appearance-based methods rely on a large number of input samples, and the results of different training samples often vary greatly, the performance is not very stable. While the feature-based methods mainly uses the local geometric features of the eyes to locate, without the need of input samples and with robust perfor-

mance. Therefore, the feature-based methods are widely used in the iris center localization, including the IDO-based method proposed by Daugman [7], the isophote-based method proposed by Valenti et al [19,20], and the gradient-based method proposed by Timm et al [18]. In the IDO-based method, an integrodifferential operator (IDO) is used for the pupil localization. The method not only has a low localization accuracy, but also has a high computation cost. However, for the isophote-based method, not only is it invariant to the rotation and linear transformation of light, but also it is more efficient. The method is also used in [14], where a shape regression method is used to locate coarsely, and then the iris center is located accurately by the isophote-based method. Furthermore, the gradient-based method is widely used at present because of its high localization accuracy. The method is also reported in [4], where FFT transform is used to enhance its efficiency. In addition, Zhang et al. [24] propose a localization method, where the two methods based on isophote and gradient are employed and integrated. Its localization accuracy is improved, but it is complicated with many not easy to set.

To sum up, we can see that the isophote-based method has the characteristic of invariance to the transformation of light and the gradient-based method has the advantage of a high localization accuracy. Following the work in [24], the above two methods are further integrated and refined in the paper.

1.3. Contributions of the paper

In order to achieve accurate iris center location in low-resolution images captured in visible light, an energy map synthesis method based on gradient [18], isophote [19,20], and midpoint the eye ROI (Region Of Interest) is proposed in this paper. The algorithm proposed is verified on three public databases. The main contributions of this paper are summarized as follows.

1) An Energy map synthesis method based on gradient [18], isophote [19,20], and midpoint of eye ROI is proposed, which has the advantages of higher localization accuracy in the image gradient method, invariance to the rotation and linear transformation of light in the isophote method, and iris center close to the midpoint of eye ROI.

2) A post-processing correction method is adopted proposed in our previous work [5], including correction for closed eyes and large deviation of iris center to further improve the localization accuracy.

3) A modeling approach of the energy map based on the eye ROI midpoint is adopted proposed in previous work [5] to provide a feasible way for energy map synthesis and localization accuracy improvement.

2. Flow chart of the proposed method

The flow chart of the iris center localization by the proposed method is shown in Figure 1.

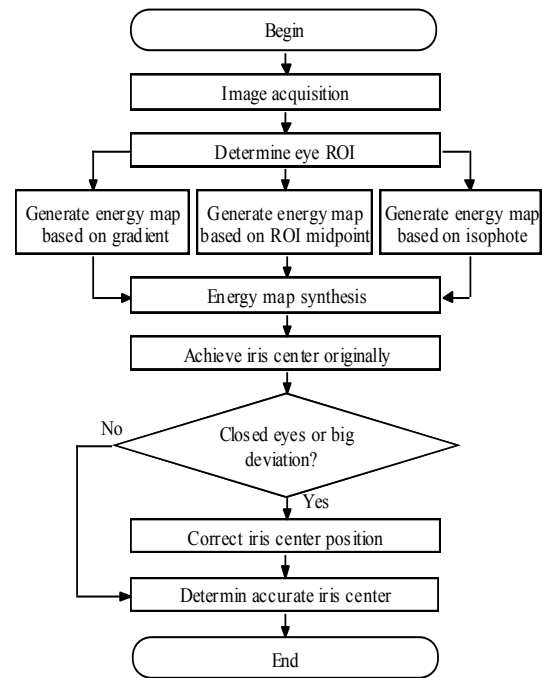


Fig. 1. Flow chart of the iris center localization by the proposed method

2.1. Energy map based on image gradient

In this paper, the basic gradient method proposed by Timm et al. [18] is improved, which mainly reflected in the weight design in order to reduce the influence of some specular reflection on the localization of iris center. The gradient-based localization method of the iris center is described below, followed by how to generate the energy map based on this method. In addition, an adaptive threshold segmentation method is introduced to improve efficiency and robustness.

2.1.1. Basic gradient method

The schematic diagram of iris center localization based on the gradient is shown in Figure 2, where the

inner circle represents the pupil, the outer circle represents the iris, c is the iris center, p_i is a point of the iris edge, d_i is the displacement vector from c to p_i , and g_i is the gradient vector of gray level at p_i .

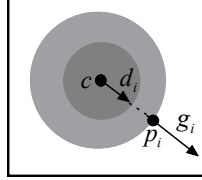


Fig. 2. Schematic diagram of iris center localization method based on gradient

Since the iris center is the darkest and its edge is lighter, if g_i and d_i are normalized into unit vectors, when the two vectors are in the same direction, their dot product is the largest. When the quadratic sum of their dot product is the largest, the center point is the iris center, which can be calculated by

$$\begin{cases} \hat{c} = \arg \max_c \left\{ \frac{1}{n} \sum_{i=1}^n (d_i^T g_i)^2 \right\} \\ d_i = \frac{p_i - c}{\|p_i - c\|_2}, \forall i: \|g_i\|_2 = 1 \end{cases} \quad (1)$$

where n is the number of all pixels in the image. In order to improve the localization accuracy of the iris center, the weighted ω_{ci} is added, so the formula of iris center localization can be rewritten as

$$\begin{cases} \hat{c} = \arg \max_c \left\{ \frac{1}{n} \sum_{i=1}^n \omega_{ci} (d_i^T g_i)^2 \right\} \\ d_i = \frac{p_i - c}{\|p_i - c\|_2}, \forall i: \|g_i\|_2 = 1 \end{cases} \quad (2)$$

Then how do we design the weight? Because the iris center is the darkest and its gray value is the smallest, the weight of a possible center point is given by

$$\omega_{ci} = 255 - I_i, \quad (3)$$

where, I_i is its gray value. Thus, the closer to the iris center the point is, the greater its weight is.

2.1.2. Improved gradient method

However, bright spots often occur or even cover the iris region because of specular reflection, which makes gray value in the region increased. If the gray value is

simply used as the weight, the error will inevitably increase. Therefore, this weight is multiplied by a coefficient k less than 1 to reduce the influence of specular reflection on the gray weight, and a small constant value m is added to enhance its robustness. Then the final weight can be expressed by [5]

$$\omega_{ci} = (255 - I_i) * k + m. \quad (4)$$

In addition, when the angle between d_i and g_i is obtuse, as shown in Figure 3, the possible iris center is located outside the iris, which is clearly unreasonable.

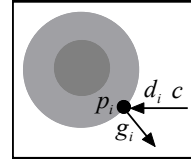


Fig. 3. The possible iris center is outside the iris

In order to eliminate the points outside the iris, the formula is further improved. The dot product of displacement vector and gradient vector is negative for the points outside the iris, so only the positive part will be retained. Then the formula of iris center localization can be rewritten as

$$\begin{cases} \hat{c} = \arg \max_c \left\{ \frac{1}{n} \sum_{i=1}^n \omega_{ci} D_i^2 \right\} \\ D_i = \max \{0, d_i^T g_i\} \\ d_i = \frac{p_i - c}{\|p_i - c\|_2}, \forall i: \|g_i\|_2 = 1 \\ \omega_{ci} = (255 - I_i) * k + m \end{cases} \quad (5)$$

2.1.3. Energy map based on the improved gradient method

In terms of the curly brace part of the top one in (5), the gray value of points in the eye ROI can be obtained, and the resulting image is the energy map based on the gradient method.

Some examples of the energy map are shown in the second column in Table 1, where the image number is shown in the first column. Furthermore, the point with the highest gray intensity in the energy map is the iris center obtained by the improved gradient method.

Table 1

Synthesis process of energy map

Image number	Energy map based on gradient method	Energy map based on isophote method	Energy map based on the eye ROI midpoint	Synthetic energy map
150				

324				
494				
576				
886				
990				
1005				
1057				
1179				
1199				

2.1.4. Adaptive threshold segmentation to improve efficiency and robustness

In order to improve the efficiency of the algorithm, the dynamic adaptive threshold segmentation method is adopted to filter the gradient. If the gradient magnitude (Mag) is greater than the threshold ($Thresh$), the gradient is divided by the magnitude to get the normalized result. Otherwise, the gradient is set to 0. Thus the gradient can be expressed as

$$gradient = \begin{cases} gradient / Mag & \text{if } (Mag > Thresh) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

What's more, the gradient threshold ($Thresh$) is set to the dynamic adaptive threshold ($stdDevFactor * stdDevMag + meanMag$), where $meanMag$ is the mean of the gradient magnitude, $stdDevMag$ is its standard deviation, and $stdDevFactor$ is the standard deviation coefficient. The gradient threshold can vary with its magnitude, so the dynamic adaptive threshold method also enhances the robustness of system.

2.2. Energy map based on isophote

Because the rotation and linear transformation of light do not affect the shape of objects, isophote has the advantage of invariance to it. Therefore, the energy map based on the isophote method is used to correct that based on the gradient in order to reduce the influence of light on iris center localization.

Isophotes refers to the curves that connect points with the same intensity in the image. Because they do not intersect with each other, the image can be described by them. In order to locate the iris center with the isophotes, we try to find a corresponding circle at each point of each isophote to best approximate the

curve. Then the curvature circle exactly satisfies this requirement, so the curvature of the isophotes is calculated. In terms of the relationship between the illumination (L) and the isophote, the curvature k of the isophote can be derived as [19]

$$k = -\frac{L_y^2 L_{xx} - 2L_x L_{xy} L_y + L_x^2 L_{yy}}{(L_x^2 + L_y^2)^{3/2}}, \quad (7)$$

where L_x and L_y are first-order partial derivation of the illumination function in x and y directions respectively; L_{xx} and L_{yy} are the second-order partial derivation in the x and y directions, respectively; and the L_{xy} is the first-order partial derivation first in the x direction and then in the y direction. In order to vote for the center of curvature, we need to calculate the curvature radius of the isophote, which is the reciprocal of the curvature of the isophote. By multiplying the radius of curvature by the direction with the largest change in illumination (namely the gradient), the displacement vector D from the estimated center point can be obtained as

$$D(x, y) = -\frac{\{L_x, L_y\}(L_x^2 + L_y^2)}{L_y^2 L_{xx} - 2L_x L_{xy} L_y + L_x^2 L_{yy}}. \quad (8)$$

If the gray intensity at the boundary of an object is the same, the isophote at the edge is the same, and the shape of the isophote is the same as that of the object. Then the displacement vector can be mapped to an accumulator, and the iris center can be obtained by voting according to the accumulated result. However, in fact, we cannot guarantee that the isophote at the edge of an object is the same. Therefore, the curvedness operator is introduced, which is defined as

$$curvedness = \sqrt{L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2}. \quad (9)$$

The curvedness indicates how much the shape of an object is bent. At the edge of the object, the curvedness

is the greatest. Therefore, the curvedness is accumulated as the voting mechanism, and the cumulative result is convolved with the Gaussian function to generate the energy map. The following three constraints should be satisfied before the curvedness accumulation.

1) Curvature is negative.

The sign of curvature of an isophote depends on the intensity at the outer boundary of the curve. The gray intensity at the outer boundary of iris is the largest, and it varies from large to small from the boundary to the center direction, with a negative gradient. Therefore, only the isophotes with negative curvature (namely negative gradient) are accumulated to vote.

2) Magnitude of the displacement vector is in a certain range.

The magnitude of the displacement vector is the curvature radius of the isophote. Because the iris radius is in a certain range, the magnitude of the displacement vector should be limited in the range.

3) Gray of the image is in a lower range.

Because the iris is darker, its gray intensity should be limited in a lower range to ensure that its center is located in its region. For the images listed on the leftmost column in the Table 1, the energy maps generated in accordance with the above isophote method are shown in the third column. Moreover, based on the method, the iris center can be obtained by finding the point with the maximum gray value in the energy map.

2.3. Energy map based on the eye ROI midpoint

Since the iris center is near the eye ROI midpoint, the energy map based on the eye ROI midpoint [5] is adopted to correct the energy map. For the clarity of the paper, the modeling process of the energy map is restated here. Since the eye ROI midpoint is very close to the iris center, the gray value of the eye ROI midpoint is modeled as the maximum in this energy map. Meanwhile, in order to ensure robustness, the gray value gradually decreases from the midpoint of the eye on ROI within a certain radius. The farther away it is from the eye ROI midpoint, the smaller the gray value is, and the gray value in the farthest position is minimum 0. Accordingly, a two-dimensional Gaussian function can be used to describe the energy map. However, Gaussian function is mainly composed of an exponent which ranges from 0 to 1, while the gray value of an image ranges from 0 to 255. Therefore, the two-dimensional Gaussian function with the peak value 255 in the eye ROI midpoint, can be selected as the

corresponding energy function, which can be modeled as [5]

$$E_m(x, y) = 255 * e^{-\frac{(x-x_m)^2 + (y-y_m)^2}{\sigma^2}}, \quad (10)$$

where (x_m, y_m) is the two-dimensional coordinate of the eye ROI midpoint, σ is the energy decay radius in the energy map. According to (10), the energy maps based on the eye ROI midpoint can be generated, shown in the fourth column in Table 1.

2.4. Energy map synthesis and iris center localization

The energy maps based on the gradient, the isophote, and the eye ROI midpoint are synthesized in terms of a certain weight to generate the final one. The image fusion function of `addWeight` in OpenCV[17] is used to implement the function. Since the function can only synthesize two energy maps, the energy maps based on the isophote and the eye ROI midpoint are first synthesized. Because the localization accuracy of the two methods is similar (see Table 7 below), their weights are set to 0.5 respectively. Then their synthetic energy map is synthesized with the energy map based on the gradient to generate the final one. The weight is chosen by experiment results of localization accuracy when e (maximum normalized error) is within 0.05 after energy map synthesis, shown in Table 2. The first column of the Table is the weight of energy map based on the gradient. When the weight is 0.1, the weight of synthetic energy map based on the isophote and the midpoint of the ROI is 0.9, that is, the two weights add up to 1. From Table 2 we can see when the weight of energy map based on the gradient is 0.6 marked in bold, the localization accuracy is the highest in the range of $e \leq 0.05$, so its weight is taken as 0.6, while the weight of synthetic energy map based on the isophote and the eye ROI midpoint is set to 0.4.

Table 2

Localization accuracy corresponding to different weight

Weight	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
0.1	84.5545%	98.0858%	100%
0.2	85.4786%	97.6238%	99.802%
0.3	86.9307%	97.6238%	99.736%
0.4	87.7888%	96.8977%	99.604%
0.5	88.3828%	96.8317%	99.3399%
0.6	88.8449%	96.3696%	99.2079%
0.7	88.5149%	95.5116%	98.7459%
0.8	86.6007%	93.7954%	98.1518%
0.9	83.1683%	92.0792%	97.0957%

For the images listed on the leftmost in the Table 1, the final synthetic energy maps are shown in the last

column. The point with the largest gray value in the synthetic energy map is the iris center determined originally.

2.5. Correction for the closed eyes and the other large Deviations

Because both the gradient-based method and the isophote-based method are suitable for the open eye state, their localization errors in the closed eye state are large. It is necessary to detect the closed eye and correct the location of the iris center. It is known that the ratio of width to height of the eye ROI will increase when the eye is closed. Therefore, whether the eyes are closed can be judged by the ratio. When the ratio is larger than a threshold, we think the person in the image is in the closed or semi-closed state. The threshold can be set properly according to monitoring the actual ratio of the width and height of the eye ROI. Because iris center is close to the eye ROI midpoint, in the closed eye state, the iris center is approximated by the eye ROI midpoint.

In addition, the other large deviations of the iris center location caused by occlusion, specular reflection, too strong light or too weak light and so on will be corrected. When the iris center obtained by energy map is far from the eye ROI midpoint and close to its edge, or the interpupillary distance is out of a certain threshold range, which indicates that the localization error is large. Then by judging which eye is close to the edge of the ROI, we can use the midpoint of its ROI to correct it. For example, if the iris center (marked by \bullet) obtained by the above method is far from the midpoint (marked by $+$) of the eye ROI (represented by rectangle), just as Figure 4, the position of the iris center will be corrected. Likewise, as shown in Figure 5, when interpupillary distance is too large in (a) and (b) or too small in (c) and (d), the correction will also be carried out. In (a) and (c), their right eyes will be corrected, while in (b) and (d), their left eyes will be corrected. Figure 4 and Figure 5 show the schematic diagram of correction for large deviations, some examples of real images are shown in Table 5 in section 4.3.

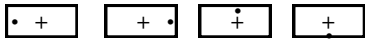


Fig. 4. Large deviation of iris center location for one eye

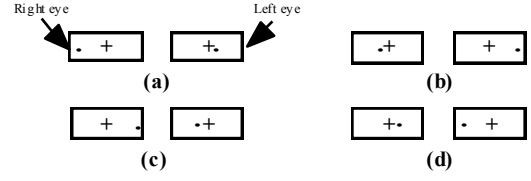


Fig. 5. Interpupillary distance is too large or too small

The adopted post-processing correction algorithm is detailed in the literature [5]. It's worth noting that the threshold parameters in the algorithm are all divided by eye width, so they are normalized. After correction for closed eyes and for other large deviations of the iris center, the final accurate iris center position is obtained.

3. Result evaluation

3.1. Evaluation index

Maximum normalization localization error [8] is used to evaluate the localization accuracy of iris center. The evaluation formula is given by

$$e = \frac{\max(\|\hat{c}_l - c_l\|_2, \|\hat{c}_r - c_r\|_2)}{\|c_l - c_r\|_2}, \quad (11)$$

where \hat{c}_l and \hat{c}_r are the estimated iris center position of left eye and right eye respectively, and c_l and c_r are their actual iris center position respectively [5]. It can be seen that formula (11) represents the ratio of the maximum absolute localization error of two eyes to their distance, which is normalized. In addition, there are also minimum and average normalization localization error. The former is the ratio of their minimum absolute localization error to the distance between them. The latter is the ratio of their average absolute localization error to the distance between them.

According to the range of localization error, localization accuracy is often divided into the following three situations: $e \leq 0.05$ indicates that the localization error is roughly within the distance from the pupil center to its boundary; $e \leq 0.10$ represents that the error is within the distance from the iris center to its boundary; $e \leq 0.25$ denotes that the error is basically within the distance from the center of eye to the its corner. The number of images in each range of localization error is counted, divided by the total number of images, to obtain the localization accuracy in each error range.

3.2. Database

The BioID public database [16] is employed to evaluate the localization results. The database contains 1521 low-resolution images captured by webcams, including images with the closed or semi-closed eyes, those captured in strong or weak light, even those with eye areas occluded by strong specular reflection, recognized as the most challenging database.

3.3. Localization results

Face detection and facial landmarks detection are carried out on 1521 images in BioID database. At the same time, OpenCV [17] and Dlib [9] are used to detect faces. Many faces in the images cannot be detected

by the former, while only 6 faces are not detected by the latter, so face detection is realized by using Dlib. Facial landmark points in 1515 face images are detected to obtain the eye ROIs. On this basis, the above localization algorithm is adopted to locate the iris center. For the images listed in Table 1 except image 990 and image 1199, the detected face regions and eye ROIs images are shown in Figure 6. Because image 990 is similar to image 1005, and image 1199 is similar to image 1179, the two images (990 and 1199) are not shown here. From these images, we can see that many people wear glasses, and they are exposed to varying illuminations. Moreover, due to the influence of specular reflection, bright spots occur or even cover the iris region, which degenerates significantly the quality of eye ROI image.

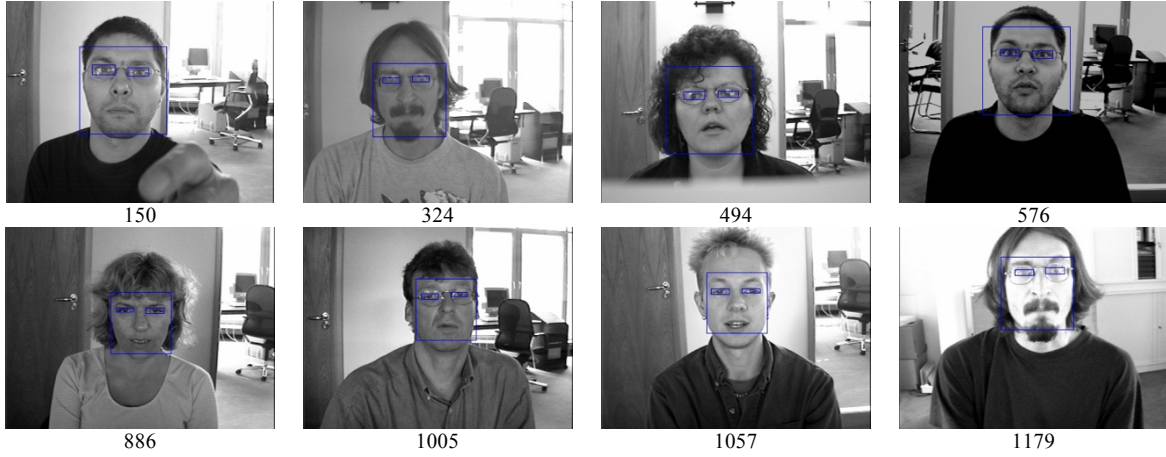


Fig. 6. Some examples of detected face regions and eye ROIs images on the BioID database

For the images listed in Table 1, the localization results before and after energy map synthesis are compared as shown in Table 3. Their localization results based on the gradient method before synthesis are shown in the second column; those based on the isophote method are listed in the third column; and those after the energy map synthesis are in the last column. The positions of the iris centers are marked by cross-lines. Furthermore, the comparison of their localization errors before and after the energy map synthesis is shown in Table 4. Due to the effects of bright spots in eye ROI, the gray value in iris region increase, which makes the localization accuracy worse for the method depending on gray feature such as the gradient-based and the isophote-based methods, as shown in the middle two columns in Table 3 and Table 4. However, because of the correctness effect of the isophote and eye ROI midpoint on the gradient, after the energy

map synthesis, the localization accuracy is greatly improved, as shown in the last column in the two tables.

After the energy map is synthesized, correction for the closed eyes and for other large deviations are carried out. Some examples of localization results before and after the correction are shown in Table 5, and the comparison of their localization errors is shown in Table 6. It can be seen that for image 291 and image 879, the correction for the closed eyes is performed. While for the four images from 325 to 588 listed in Table 5, large deviations caused by poor image quality under dark light are also corrected. Furthermore, for the remaining four images, large deviations caused by bright illumination and specular reflection are corrected. The results show that post-processing correction can effectively refine localization accuracy in further.

Table 3

Comparison of localization results (LR) before and after the synthesis of energy map

Image number	LR based on gradient	LR based on isophote	LR after energy map synthesis
150			
324			
494			
576			
886			
990			
1005			
1057			
1179			
1199			

Table 4

Comparison of localization errors (LE) before and after synthesis of energy maps

Image number	LE based on gradient	LE based on isophote	LE after energy map synthesis
150	0.278068	0.246942	0.0282335
324	0.253837	0.0507673	0.0227038
494	0.0287658	0.231917	0.0203405
576	0.0212334	0.192276	0
886	0.191959	0.047619	0.0238095
990	0.051778	0.301915	0.0231558
1005	0.050702	0.137924	0.0320668
1057	0.078363	0.0896116	0.021734
1179	0.208965	0.257192	0.0415855
1199	0.0702555	0.274806	0.0222167

Table 5

Comparison of localization results (LR) before and after correction

Image number	LR before correction	LR after correction
157		
291		
325		
361		
409		
588		
652		
769		
776		
879		

Table 6

Comparison of localization errors (LE) before and after correction

Image number	LE before correction	LE after correction
157	0.201665	0.0425147
291	0.154212	0.0344828
325	0.220797	0.048766
361	0.282542	0.0485987
409	0.200378	0.021734
588	0.254099	0.0321412
652	0.188654	0.0416667
769	0.176777	0.0222167

776	0.148316	0.0344828
879	0.103975	0.0519875

Based on the different method, the localization accuracy in each error range is shown in Table 7, where e is the maximum normalization localization error, computed by formula (11). The corresponding localization accuracy is the proportion of the number of images within each localization error to the total number of images.

Table 7

Iris center localization accuracy on the BioID database

Method	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
Gradient method	81.1%	91.5%	97.4%
Isophote method	63.8%	77.1%	96.8%
Midpoint of ROI	68.0%	98.0%	99.9%
Energy map synthesis	88.8%	96.4%	99.2%
After correction	90.1%	98.8%	99.9%

The localization accuracy based on the gradient method is shown in the first row data of the Table 7. The second row is the localization accuracy based on the isophote. The third row is that using the eye ROI midpoint as the iris center. The fourth row is that after the energy map synthesis. The last row shows that after correction for the closed eyes and for other large deviations. Compared with the gradient-based method before the synthesis, the localization accuracy after the energy map synthesis is improved by 7.7% in the range of $e \leq 0.05$, 4.9% within $e \leq 0.1$, and 1.8% within $e \leq 0.25$. It can be seen that the energy map synthesis has a significantly effect on improving the localization accuracy, especially in the small range of localization error. After correction for the closed eyes and other large deviations, the localization accuracy is increased by another 1.3% within $e \leq 0.05$, 2.4% within $e \leq 0.1$, and 0.7% within $e \leq 0.25$. It is obvious that the localization accuracy is further improved.

The above results are based on the maximum normalization localization error. Similarly, the localization accuracy based on the average and minimum normalized localization error is also calculated. That based on the maximum, average and minimum normalized localization error is shown in Table 8, and their trend curves are shown in Figure 7.

Table 8

Localization accuracy based on maximum, average and minimum normalization error

Measurement of localization accuracy	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
Maximum normalization error	90.1%	98.8%	99.9%
Average normalization error	96.0%	99.5%	99.9%
Minimum normalization error	98.7%	99.9%	99.9%

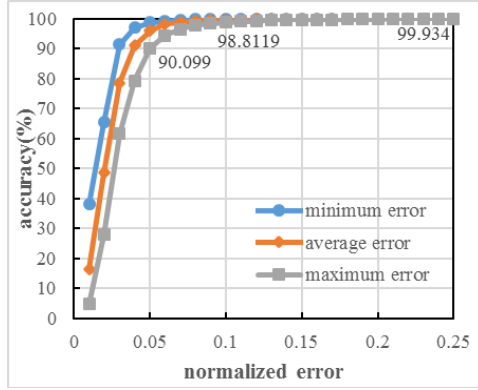


Fig. 7. Localization accuracy curve

It is clear that the localization accuracy based on the average and minimum normalization error is higher than that based on the maximum one. In particular, the proportion of images with minimum normalization error in the range of 0.05 has amounted to 98.7%, which proves that the localization accuracy of the proposed method is very high. Compared with current advanced methods on the BioID database, as shown in Table 9.

Table 9

Comparison of localization accuracy on the BioID database

Method	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
Timm2011[18]	82.5%	93.4%	98.0%
Valenti2008[19]	84.1%	90.9%	98.5%
Valenti2012[20]	86.1%	91.7%	97.9%
Markus2014[12]	89.9%	97.1%	99.7%
Ren2014[15]	77.1%	92.3%	99.0%
Cai2015[4]	84.1%	95.6%	99.8%
Pang2015[14]	83.6%	96.1%	98.6%
Zhang2016[24]	85.7%	93.7%	99.2%
Araujo2017[3]	88.3%	92.7%	98.9%
Xia2018[23]	87.1%	98.7%	99.9%
Ahmed 2019[1]	87.0%	95.2%	99.0%
Laddi2019[10]	81.4%	92.2%	97.5%
Proposed method	90.1%	98.8%	99.9%

The localization accuracy of the proposed method is shown in the last row of the table, marked in bold. It is obvious the localization accuracy within each error range is the highest, compared the other methods listed in the table. It can be seen that the localization accuracy has outperformed the state-of-the-art methods. Although the localization accuracy based on the gradient method in this paper are not as high as that obtained by Timm, the final localization accuracy by the proposed method is higher. If the localization accuracy based on the gradient method is improved, the final localization accuracy will continue to rise.

4. Generalization ability and robustness

In order to verify the generalization ability and robustness of the proposed method, the algorithm is tested on the other two datasets which are Talking Face Video database [22] and the MUCT Face database [6]. The Talking Face Video database includes 5000 images with various pose and expressions faces, while the MUCT Face database contains 3755 images with different pose people in varying illuminations. Among these images, only the faces in three images on the MUCT Face database could not be detected, and the remaining faces were all detected. The following results are based on the detected face images. Some examples of detected face regions, eye ROIs (marked with rectangles) and corresponding localization result (marked with cross lines) images on the Talking Face Video database [22] and the MUCT Face database [6] are shown in Figure 8 and Figure 9, respectively.



Fig. 8. Examples of localization result images on Talking Face Video database

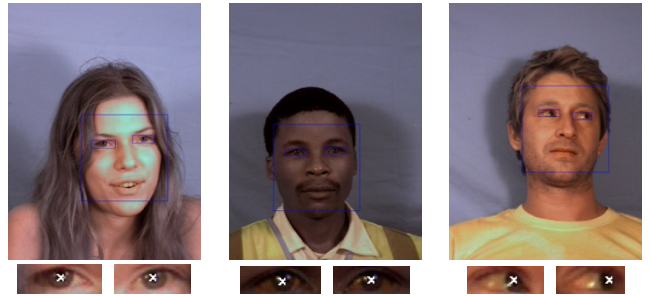




Fig. 9. Examples of localization result images on the MUCT Face database

The localization accuracy of iris center on the two databases is shown in Table 10 and Table 11 respectively. The first row data shown in the two tables is the localization accuracy based on the gradient method, the second row is that based on the isophote method, the third row is that based the energy map synthesis, the last row is that after post-processing correction. Comparing the first row with the third row in Table 10, we can see that when the localization error is within 0.05, the localization accuracy is improved by 11.7%; when the localization error is within 0.1, the localization accuracy is increased by 2.1%. It can be seen that the localization accuracy is improved dramatically by the proposed energy map synthesis method, compared with the gradient method. Comparing the last two rows, we can see that after post-processing correction, when the localization error is within 0.05, the localization accuracy is increased by another 0.4%; within 0.1, improved by another 1.0%. It can be seen that the localization accuracy is improved in further by the post-processing correction. Furthermore, from Table 11, we can find the interesting thing that though the localization accuracy based on the isophote method is very low (only 32.6% and 51.6% respectively) on the MUCT Face database when the localization error is within 0.05 and within 0.1, the localization accuracy by the energy map synthesis amounts to 83.8% and 96.7% respectively. Compared with the gradient method, the localization accuracy by the energy map synthesis is improved by 13.5% within 0.05 localization error. The results validate the effectiveness of the proposed energy map synthesis method. After post-processing correction, the localization accuracy is refined by another 1.6% in further within 0.05 localization error. The results show the post-processing correction is also very effective.

Table 10

Iris center localization accuracy on Talking Face Video database

Method	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
Gradient method	81.1%	96.6%	99.9%
Isophote method	69.8%	87.7%	96.8%
Energy map synthesis	92.8%	98.7%	99.9%
After correction	93.2%	99.7%	99.96%

Table 11

Iris center localization accuracy on the MUCT Face database

Method	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
Gradient method	70.3%	89.7%	98.5%
Isophote method	32.6%	51.6%	96.6%
Energy map synthesis	83.8%	96.7%	99.7%
After correction	85.4%	97.6%	99.9%

Moreover, compared with other advanced iris center localization methods on Talking Face Video and the MUCT Face databases, the results of localization accuracy are shown in Table 12 and Table 13 respectively. It is obvious that the localization accuracy by the proposed method on the two databases has outperformed the state-of-the-art approaches listed in the two tables. It follows that the algorithm has strong robustness and good generalization ability.

Table 12

Comparison of localization accuracy on Talking Face Video database

Method	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
Pang2015[14]	--	96.2%	--
Ahmed2019[1]	--	98.8%	--
Laddi2019[10]	90.2%	98.7%	99.5%
Proposed method	93.2%	99.7%	99.96%

Table 13

Comparison of localization accuracy on the MUCT face database

Method	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
Laddi2019[10]	78.6%	88.3%	92.2%
Proposed method	85.4%	97.6%	99.9%

Furthermore, the new proposed method is compared with our previously proposed method in [5]. The main difference between them consists in the energy map based on the isophote proposed in the new method. The energy map based on the isophote is used to correct the energy map in order to acquire a more accurate position of the iris center. Because the isophote-based method has the advantage of the invariance to the rotation and linear transformation of light, its correctness effectiveness in the new proposed method makes the localization accuracy improved with varying illuminations. For databases with varying illuminations such as BioID and MUCT, the localization accuracy of the new proposed method using the isophote, is higher than the previous one [5], shown in the first two rows and the last two rows of Table 14. While for the database with invariant illu-

minations such as Talking Face, the localization accuracy of the new method is lower than the previous one, shown in the third and fourth rows of Table 14. It is because that the isophote method has no merit for invariant illuminations, and the localization accuracy of the method (shown in the second row of Table 10) is low. It follows that the new proposed method is suitable for varying illuminations.

Table 14

Comparison of localization accuracy on the three face databases

Method	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.25$
Dai2020 [5] (BioID)	89.6%	98.7%	99.9%
Proposed method (BioID)	90.1%	98.8%	99.9%
Dai2020 [5] (Talking Face)	94.3%	99.8%	99.98%
Proposed method (Talking Face)	93.2%	99.7%	99.96%
Dai2020 [5] (MUCT)	84.6%	97.2%	99.9%
Proposed method (MUCT)	85.4%	97.6%	99.9%

5. Conclusion and further work

In order to accurately locate the iris center in the low-resolution face images collected by low-cost cameras in visible light, we propose an energy map synthesis method, which synthesizes the energy maps based on the gradient, the isophote, and the eye ROI midpoint. The algorithm is tested on the challenging BioID database. Compared with the method based on the gradient, the localization accuracy after energy map synthesis is improved significantly whether in the smaller range of maximum normalization localization error ($e \leq 0.05$) or in the larger range of error ($e \leq 0.1$), thereby comprehensively improving the localization accuracy. Furthermore, the post-processing correction is carried out to increase the localization accuracy in further. The localization accuracy by the proposed method is up to 90.1% when $e \leq 0.05$, 98.8% when $e \leq 0.1$, which has outperformed the listed state-of-the-art methods.

Furthermore, the algorithm is verified on the other two databases, which are Talking Face Video database with images of various pose and expression, and the MUCT Face database with images of various pose and varying illuminations. The results show that the proposed method has a strong robustness to the pose and illumination, and it has good effectiveness and generalization. Especially, compared with the energy map synthesis method based on the gradient and the eye ROI midpoint [5], the proposed method in the paper is refined with the energy map based on the isophote, which makes the method more suitable to varying illuminations, because the isophote has the advantage of

the invariance to the rotation and linear transformation of light.

In addition, since the two methods based on the gradient and the isophote are adopted to locate the iris center, the operation efficiency of the program is lower than one method. Therefore, we will optimize the program to decrease the consumed time in the future. Furthermore, though the final localization accuracy by the proposed method on the BioID database (shown in the last row in Table 9) is much higher than that by the gradient-based method proposed by Timm (shown in the first row data in Table 9), the localization accuracy based on the gradient method in this paper (shown in the first row data in Table 7) are not as high as that obtained by Timm. Therefore, for the proposed algorithm, there is still room for further improvement of localization accuracy. At the same time, based on the iris center location, gaze tracking will be realized.

Acknowledgments

This work was supported in part by National Key R&D Program of China (Grant No. 2018YFB1304600), the Natural Science Foundation of China (Grant No. 51775541), CAS Interdisciplinary Innovation Team (Grant No. JCTD-2018-11), and the Natural Science Foundation of China under Grant 51575412.

References

- [1] M. Ahmed and R.H. Laskar, Eye center localization in a facial image based on geometric shapes of iris and eyelid under natural variability, *Image and Vision Computing* **88** (2019), 52-66.
- [2] S.M. Anzalone, J. Xavier, S. Boucenna, L. Billeci, A. Narzisi, F. Muratori, D. Cohen, and M. Chetouani, Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment, *Pattern Recognition Letters* **118** (2019), 42-50.
- [3] G.M. Araujo, F.M.L. Ribeiro, W.S.S. Júnior, E.A.B.d. Silva, and S.K. Goldenstein, Weak Classifier for Density Estimation in Eye Localization and Tracking, *IEEE Transactions on Image Processing* **26** (2017), 3410-3424.
- [4] H. Cai, H. Yu, C. Yao, S. Chen, and H. Liu, Convolution-based means of gradient for fast eye center localization, in: *2015 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2015, pp. 759-764.
- [5] L. Dai, J. Liu, Z. Ju, and Y. Gao, Iris Center Localization Using Energy Map With Image Inpaint Technology and Post-Processing Correction, *IEEE Access* **8** (2020), 16965-16978.
- [6] T.M.F. Database. The MUCT Face Database. Accessed: Mar. 20, 2019. [Online]. Available: <http://www.milbo.org/muct/>
- [7] J.G. Daugman, HIGH CONFIDENCE VISUAL RECOGNITION OF PERSONS BY A TEST OF

- STATISTICAL INDEPENDENCE, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15** (1993), 1148-1161.
- [8] O. Jesorsky, K.J. Kirchberg, and R.W. Frischholz, Robust Face Detection Using the Hausdorff Distance, in: *Audio- and Video-Based Biometric Person Authentication*, J. Bigun and F. Smeraldi, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, pp. 90-95.
- [9] D.E. King, Dlib-ml: A Machine Learning Toolkit, *Journal of Machine Learning Research* (2009), 1755-1758.
- [10] A. Laddi and N.R. Prakash, Eye gaze tracking based directional control interface for interactive applications, *Multimedia Tools and Applications* (2019).
- [11] J. Lee, H. Park, S. Lee, T. Kim, and J. Choi, Design and Implementation of an Augmented Reality System Using Gaze Interaction, in: *2011 International Conference on Information Science and Applications*, 2011, pp. 1-8.
- [12] N. Markuš, M. Frliak, I.S. Pandžić, J. Ahlberg, and R. Forchheimer, Eye pupil localization with an ensemble of randomized trees, *Pattern Recognition* **47** (2014), 578-587.
- [13] O. Palinko, F. Rea, G. Sandini, and A. Sciutti, Eye gaze tracking for a humanoid robot, in: *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, 2015, pp. 318-324.
- [14] Z.Y. Pang, C.S. Wei, D.D. Teng, D.H. Chen, and H.Z. Tan, Robust Eye Center Localization through Face Alignment and Invariant Isocentric Patterns, *Plos One* **10** (2015), 19.
- [15] Y. Ren, S. Wang, B. Hou, and J. Ma, A Novel Eye Localization Method With Rotation Invariance, *IEEE Transactions on Image Processing* **23** (2014), 226-239.
- [16] B.T. Research. The BioID Face Database. Accessed: Dec. 22, 2018. [Online]. Available: <https://www.bioid.com/facedb/>
- [17] O. team. OpenCV. Accessed: Jul. 17, 2018. [Online]. Available: <https://opencv.org/>
- [18] F. Timm and E. Barth, ACCURATE EYE CENTRE LOCALISATION BY MEANS OF GRADIENTS, in: *Visapp 2011: Proceedings of the International Conference on Computer Vision Theory and Applications*, L. Mestetskiy and J. Braz, eds., 2011, pp. 125-130.
- [19] R. Valenti and T. Gevers, Accurate eye center location and tracking using isophote curvature, in: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1452-1459.
- [20] R. Valenti and T. Gevers, Accurate Eye Center Location through Invariant Isocentric Patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34** (2012), 1785-1798.
- [21] F. Vicente, Z. Huang, X. Xiong, F.D.I. Torre, W. Zhang, and D. Levi, Driver Gaze Tracking and Eyes Off the Road Detection System, *IEEE Transactions on Intelligent Transportation Systems* **16** (2015), 2014-2027.
- [22] T.F. Video. Talking Face Video database. Accessed: Mar. 20, 2019. [Online]. Available: http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html
- [23] Y. Xia, J. Lou, J. Dong, G. Li, and H. Yu, SDM-Based Means of Gradient for Eye Center Localization, in: *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2018, pp. 862-867.
- [24] W. Zhang, M. Smith, L. Smith, and A. Farooq, Eye center localization and gaze gesture recognition for human-computer interaction, in: 2016, p. 314.
- [25] Z.W. Zhu and Q. Ji, Eye and gaze tracking for interactive graphic display, *Machine Vision and Applications* **15** (2004), 139-148.